INVDOCK: A Method and Software for Computer Automated Prediction of Protein Targets of Small Molecules

Chen Yu Zong

- Purpose.
- Background Info.
- Description of Technology.
- Performance Analysis

Purpose

- To provide a new method for low-cost and high-speed prediction of protein and nucleic acid targets of a small molecule.

- Potential applications:
  1. Identification of unknown and secondary therapeutic targets of drugs, drug leads, drug candidates, natural products, etc.
  2. Prediction of drug targets related to side effect and toxicity (drug safety evaluation).
  3. Prediction of targets related to drug ADME (pharmacokinetics analysis).
  4. Identification of unknown receptors of a ligand (pathway analysis).
Background Info

Why study protein targets of a molecule?

Therapeutic Targets


Background Info

Why study protein targets?

Prediction of side effect, toxicity, pharmacokinetics and pharmacogenetics

MITOCHONDRIAL TARGETS OF DRUG TOXICITY

K. B. Wallace and A. A. Starkov
Department of Biochemistry and Molecular Biology, University of Minnesota School of Medicine, Duluth, Minnesota 55812; e-mail: kwallace@umn.edu, astarkov@umn.edu

Key Words oxidative phosphorylation, uncouplers, bioenergetics, permeability transition, redox cycling

Abstract Mitochondria have long been recognized as the generators of energy for the cell. Like any other power source, however, mitochondria are highly vulnerable to inhibition or uncoupling of the energy harnessing process and as a high risk for catastrophic damage to the cell. The exquisite structural and functional characteristics of mitochondria provide a number of primary targets for sensitizer-induced bioenergetic failure. They also provide opportunities for selective delivery of drugs to the mitochondria. In light of the large number of natural, commercial, pharmaceutical, and environmental chemicals that manifest their toxicity by interfering with mitochondrial bioenergetics, it is important to understand the underlying mechanisms. The significance is further underscored by the recent identification of bioenergetic control points for cell replication and differentiation and the realization that mitochondria play a determinate role in cell signaling and apoptotic modes of cell death.

1997, 37:269-296

MOLECULAR MECHANISMS OF GENETIC POLYMORPHISMS OF DRUG METABOLISM

Urs A. Meyer
Bischoff Institute of the University of Basel, CH-4058 Basel, Switzerland

Ulrich M. Zanger
Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, D-50354 Stuttgart, Germany

KEY WORDS: Genetic polymorphism, CYP2C9, CYP2C19, N-acetyltransferase, drug metabolism, pharmacogenetics

ABSTRACT

One of the major causes of individual variation in drug responses is genetic variations of drug metabolizing enzymes. Genetic polymorphisms of drug-metabolizing enzymes give rise to distinct subgroups in the population that differ in their ability to perform certain drug biotransformation reactions. Polymorphisms are generated by mutations in the genes for these enzymes, which can be decreased, increased, or abnormally expressed or activity by multiple molecular mechanisms. Moreover, the variant alleles exist in the population at relatively high frequencies. Genetic polymorphisms have been described for most drug metabolizing enzymes. The molecular mechanisms of most polymorphisms are reviewed here.
**Background Info**

**Detection of side effect and toxicity in early stages of drug discovery**

### Significance:
- Most drug candidates fail to reach market (> 99%).
- Side effect and toxicity is an important reason (in 30-40% cases).
- Large portion of money ($350 million per drug) and time (6-12 years for a drug) has been wasted on failed drugs.

*Drug Discov Today 1997; 2:72*

**Background Info**

**Why study protein targets of a molecule?**

### Drugs from Natural Products

- From natural products to therapeutic drugs

*TIPS, May 1999, 20:190*

- Screening of bioactive compounds
- Molecular mechanism
- Further development
Background Info

**Drug targets and new drug discovery**

**Drugs from Traditional Medicine**

- Mixture of multiple herbs etc.
- Maintaining and restoring balance.
- Mutual accentuation, mutual enhancement, mutual counteraction, mutual suppression, mutual antagonism, mutual incompatibility.

- Multiple targets: therapeutic effects, symptom treatment, toxicity modulation, drug delivery, harmonization.

*Pharmacology & Therapeutics 2000, 86:191-198*

---

**Background Info**

The need for a new target-prediction method:

- Existing experimental methods costly and time-consuming.
- Limited resources: difficulty in compound synthesis and bioassay.
- Existing computer methods not designed for target searching. Inability in cavity identification and in docking to large cavities.
Background Info

Feasibility:

Protein sampling:
- Database: >15,000 3D structures in PDB.
- Protein diversity: 17% in PDB with unique sequence.
- Development of structural genomics: 10,000 unique proteins within 5 years.

Method Accuracy:
- Ligand-protein docking algorithms capable of finding binding conformations.

Computer capability:
- Increasing performance (docking of 100,000 compounds in days).
- Decreasing cost (Linux PC, Multi-processor Machine)

Technology Description

Strategy

Existing Methods:
Given a Protein, Find Putative Binding Ligands From a Chemical Database

<table>
<thead>
<tr>
<th>Compound Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 1</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Compound n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
</tr>
</thead>
</table>

| Successfully Docked Compounds as Putative Ligands |

New Method:
Given a Ligand, Find Putative Protein Targets From a Protein Database

<table>
<thead>
<tr>
<th>Protein Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein 1</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Protein n</td>
</tr>
</tbody>
</table>

| Ligand           |

| Successfully Docked Proteins as Putative Targets |

Science 1992;257: 1078
Proteins 2001;43: 217
**Technology Description**

**Key technology:**

- Ligand-protein inverse docking strategy for target identification through structure database search.
- Flexible ligand-protein docking algorithm with no restriction on cavity size and no knowledge about specific binding region within a cavity.
- Method for automated detection of all cavities in a protein or a nucleic acid.
- Development of a biomolecular cavity database for all protein and nucleic acid entries in PDB.

*Proteins* 2001;43: 217

**INVDOCK procedure:**

1. Vector-based docking of a ligand to a cavity
2. Limited conformation optimization on the ligand and side chain of biomolecule
3. Energy minimization for all atom in the binding site
4. Docking evaluation by molecular mechanics energy functions and comparison with other ligands

Successfully Docked Proteins and Nucleic Acids as Putative Targets of a Ligand

Potential Applications:

- Protein function, Proteomics, Ligand transport, Metabolism
- Therapeutic Targets, Side-Effects, Metabolism, Toxicity Function in Pathways
**Technology Description**

**INVDOCK Cavity Models**

The docked (blue) and crystal (yellow) structure of ligands in some PDB ligand-protein complexes. The PDB Id of each structure is shown.

**INVDOCK Performance Analysis**

The docked (blue) and crystal (yellow) structure of ligands in some PDB ligand-protein complexes. The PDB Id of each structure is shown.
### INVDOCK Performance Analysis

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Docked Protein</th>
<th>PDB Id</th>
<th>RMSD</th>
<th>Description of Docking Quality</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indinavir</td>
<td>HIV-1 Protease</td>
<td>1hsg</td>
<td>1.38</td>
<td>Match</td>
<td>-70.25</td>
</tr>
<tr>
<td>Xk263 Of Dupont Merck</td>
<td>HIV-1 Protease</td>
<td>1hvr</td>
<td>2.05</td>
<td>Match</td>
<td>-58.07</td>
</tr>
<tr>
<td>Vac</td>
<td>HIV-1 Protease</td>
<td>4phv</td>
<td>0.80</td>
<td>Match</td>
<td>-88.46</td>
</tr>
<tr>
<td>Folate</td>
<td>Dihydrofolate Reductase</td>
<td>1dhf</td>
<td>2.41</td>
<td>One end match, the other in slightly different orientation</td>
<td>-63.92</td>
</tr>
<tr>
<td>5-Deazaflolate Dihydrofolate Reductase</td>
<td>2dhf</td>
<td>1.48</td>
<td>Match</td>
<td>-65.49</td>
<td></td>
</tr>
<tr>
<td>Estrogen</td>
<td>Estrogen Receptor</td>
<td>1a52</td>
<td>1.30</td>
<td>Match</td>
<td>-45.86</td>
</tr>
<tr>
<td>4-Hydroxytamoxifen</td>
<td>Estrogen Receptor</td>
<td>3ert</td>
<td>0.97</td>
<td>Match</td>
<td>-55.15</td>
</tr>
<tr>
<td>Guanosine-5’-[B,G-Methylene] Triphosphate</td>
<td>H-Ras P21</td>
<td>121p</td>
<td>0.94</td>
<td>Match</td>
<td>-60.20</td>
</tr>
<tr>
<td>Glycyrrhizic Acid</td>
<td>Carboxypeptidase A α</td>
<td>3cpa</td>
<td>2.19</td>
<td>Match</td>
<td>-44.84</td>
</tr>
</tbody>
</table>

*Proteins 2001;43: 217*

### Compound Summary

<table>
<thead>
<tr>
<th>Compound</th>
<th>Potential Targets Identified</th>
<th>Experimentally Confirmed</th>
<th>Experimentally Implicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>4H-Tamoxifen</td>
<td>17</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Aspirin</td>
<td>52</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>46</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Vitamin E</td>
<td>26</td>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>

*Proteins 2001;43: 217*
### INVDOCK Performance Analysis

<table>
<thead>
<tr>
<th>Compound</th>
<th>Number of experimentally confirmed or implicated toxicity targets</th>
<th>Number of toxicity targets predicted by INVDOCK</th>
<th>Number of toxicity targets missed by INVDOCK</th>
<th>Number of toxicity targets without 3D structure or involving covalent bond</th>
<th>Number of INVDOCK predicted toxicity targets without experimental finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>15</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Indinavir</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Neomycin</td>
<td>14</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Penicillin G</td>
<td>7</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>38</td>
<td>5</td>
<td>25</td>
<td>29</td>
</tr>
</tbody>
</table>

*J. Mol. Graph. Mod 2001;20: 199*

### INVDOCK Performance Analysis

<table>
<thead>
<tr>
<th>Natural Product Drug from Chinese Medicinal Plants</th>
<th>Number of Identified Therapeutic Targets</th>
<th>Number of Confirmed or Implicated Therapeutic Targets by experiment</th>
<th>Number of Identified Toxicity/Side effect Targets</th>
<th>Number of Confirmed or Implicated Toxicity/Side Effect Targets by experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acromycine</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Allicin</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Baicalin</td>
<td>14</td>
<td>4</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Catechin</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>Camptothecine</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Dicoumarin</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Emodin</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Genistin</td>
<td>22</td>
<td>7</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>

**Conclusions**

- Ligand-protein inverse docking is useful in probing potential targets of a molecule (~50% accuracy).
- Potential application in drug therapeutic target identification, safety evaluation, and pharmacokinetics prediction for drugs, drug leads, and natural products.
- Application potential increases along with advances in structural and functional genomics.